DOCUMENT RESUME

ED 167 633 TM 008 470

AUTHOR Kidder, Steven J.
TITLE Comparing Rasch Calibration of Passage Reading Difficulty with Spache or Dale-Chall Passage Readability Estimates.
PUB DATE Apr 77
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cloze Procedure; Complexity Level; Elementary Secondary Education; *Readability; *Readability Formulas; Reading Ability; *Reading Level; Test Interpretation
IDENTIFIERS Dale Chall Readability Formula; *Rasch Model; Rasch Scaled Scores; Spache Readability Formula

ABSTRACT

The utility and precision of scaling 200 relatively short reading passages were analyzed using a Rasch-based measurement technique as compared to readability estimates using the Spache or Dale-Chall formulas. In addition, a regression model was specified for predicting the grade level of a text a student could literally comprehend based on performance using the Rasch-scaled passages. The Rasch-based scaling required approximately 10,000 students distributed evenly across grades 1 to 12. Correlation, linear, and non-linear regression techniques were used for analysis. These analyses supported the conclusion that Rasch-based scaling of passage difficulty was more precise than the scaling based on Spache or Dale-Chall readability formulas. In addition, Rasch passage difficulty could be estimated reasonably well from the Spache or Dale-Chall readability estimates, and Rasch passage difficulty could be used to accurately predict the grade level of text a student could literally comprehend. It was recommended that the 200 Rasch-scaled passages be used for monitoring increases in student ability to literally comprehend written discourse and that they be used as the dependent variable in studies or computer programs designed to predict the difficulty or readability of written discourse. A sample item is appended. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED167633

Comparing Rasch Calibration of Passage Reading Difficulty with Spache or Dale-Chall Passage Readability Estimates

Steven J. Kidder

New York State Education Department

Albany, New York 12234

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Steven J. Kidder

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM008 470

Paper presented at the meeting of the American Educational Research Association, New York, New York, April 1977.

Abstract

This study analyzed the utility and precision of scaling 200 relatively short reading passages using a Rasch-based measurement technique as compared to readability estimates using the Spache or Dale-Chall formulas. These passages were taken from the Test Development Notebook developed by the Bureau of School and Cultural Research, Division of Research, New York State Education Department. In addition, a regression model was specified for predicting the grade level of a text a student could literally comprehend based on performance using the Rasch-scaled passages. The Rasch-based scaling required approximately 10,000 students distributed evenly across grades 1 to 12. Correlation, linear, and non-linear regression techniques were used for analysis. These analyses supported the conclusion that Rasch-based scaling of passage difficulty was more precise than the scaling based on Spache or Dale-Chall readability formulas. In addition, Rasch passage difficulty could be estimated reasonably well from the Spache or Dale-Chall readability estimates, and Rasch passage difficulty could be used to accurately predict the grade level of text a student could literally comprehend. It was recommended that the 200 Rasch-scaled passages be used for monitoring increases in student ability to literally comprehend written discourse and that they be used as the dependent variable in studies or computer programs designed to predict the difficulty or readability of written discourse.

Comparing Rasch Calibration of Passage Reading Difficulty

with Spache or Dale-Chall Passage Readability Estimates

This research was designed to Rasch-calibrate the difficulty of 200 reading passages from a recently developed test of literal comprehension (O'Reilly and Streeter, 1977) and to relate these "person-free" estimates of passage difficulty to readability estimates based on the Spache (1960) or Dale-Chall (1948) formulas.[1] Of concern was the utility and precision of Rasch-based scaling as opposed to the readability scaling. Intuitively, both scaling procedures seemed to be "person-free." With regard to Rasch procedures, "test calibrations are independent of the sample of persons used to estimate item parameters" (Wright and Panchapakesan, 1969, p. 23), while readability estimates are simply linear functions of the percentage of hard words in a passage and the average sentence length of a passage.

It is commonly known that readability formulas were developed to predict the reading difficulty of various written materials (Spache, 1960; Dale and Chall, 1948). Both the Spache and Dale-Chall formulas are based on average sentence length and percentage of hard words in a sample of written discourse. The readability of written materials is reported in grade-equivalent scores by the Spache formula up to grade 3.9. The probable error of estimate "in predicting the grade level of a book by this method is 3.3 months." (Spache, 1960,

p. 118). For more difficult materials the Dale-Chall formula must be used. The Dale-Chall formula score is interpreted in terms of corrected grade level, a range of formula scores being equated with two grades. For example, a formula score from 6.0 to 6.9 means 7th or 8th grade reading material.

However, readability estimates are usually based on one or more 100-word samples from written text; readability estimates are usually not used to scale short reading passages on one continuous scale of difficulty. Nevertheless, even for short texts or passages, readability estimates seem to be the most reliable and convenient means available for scaling reading difficulty. However, when the number of words in a passage or selection drops below 100, are the estimates of readability precise? Or, if 200 to 1,500 short passages must be calibrated on the same difficulty scale, do readability formulas provide useful discriminations among passages? Having several hundred reading passages that reliably and validly measure literal comprehension (Schuder, Kidder, and O'Reilly, Note 1) and that are precisely calibrated on one continuous scale of difficulty would be a valuable evaluation resource, primarily because different test forms could easily be assembled from the scaled passages and used for survey testing, achievement monitoring over time, or tailored testing (Salter, Note 2). Readability formulas can provide initial estimates of reading difficulty; however, Rasch scaling procedures are sample-free and provide the needed precision for multiple-evaluation purposes. In fact, performance on precisely-scaled reading

passages could be used to place students in texts of appropriate difficulty.

The procedures used in deriving Rasch passage difficulties were based on the work of Rasch (1960, 1966 a, 1966 b), Wright and Panchapakesan (1969), and Wright and Mead (Note 3). The Rasch measurement model provides sample-free item analysis.

> The procedure for sample-free item analysis is based on a very simple model . . . for what happens when any person encounters any item. The [Rasch] model says that the outcome of such an encounter is governed by the product of the ability of the person and the easiness of the item and nothing more! The more able the person, the better his chances for success with any item. The more easy the item, the more likely any person is to solve it. (Wright and Panchapakesan, 1969, p. 23.)

As Wright and Mead (Note 3) point out, "application of the Rasch model gives person-free item calibrations and item-free person measurements" when responses to tests approximate the following conditions:

1. the variable measured is <u>unidimensional</u>.
2. there are no strong relationships among persons or items other than those specified by the model so that responses of persons to items are stochastically <u>independent</u> given their parameters in the model.

> 3. items and persons do not differ substantially with respect to other possible response factors not represented in the model such as item discrimination, person sensitivity, guessing or indifference. (Wright and Mead, Note 3, p. 2)

The multiple-choice cloze tests used in the present investigation provide response data that fit the Rasch basic measurement model (Hayford, Schuder, and Kidder, Note 4).

Item calibration using the Rasch model, as programmed in CALFIT by Wright and Mead (Note 3), provides Rasch item difficulties on a scale with equal-interval properties. With simple statistical techniques, these Rasch item difficulties can be "linked" together to form a continuous scale from the least to the most difficult item. In the present study, Rasch item difficulties were "linked" to span a difficulty range comparable to student abilities in grades 1 through 12. In addition, the item difficulties within each multiple-choice cloze passage were averaged, becoming Rasch passage difficulty. This resulted in one scale for 200 reading passages rank ordered from least to most difficult.

In essence, this research was designed to answer the following questions: (a) Could 200 relatively short reading test passages be Rasch-calibrated on one scale of increasing difficulty? (b) Would Rasch-based scaling be more precise or sensitive than readability techniques? (c) Could Rasch passage difficulty be predicted from readability estimates? (d) Could student performance on the Rasch-scaled passages be used to accurately predict the grade level of a text the student could literally comprehend?

Scaling Precision

Method

The Rasch-based scaling procedure required a comprehensive test design and administration plan to obtain the necessary data for calibrating 200 reading passages on the same difficulty scale. This plan included the construction of a series of 75 "overlapping" test forms.[2] Each test form had two identical passages in common with the next form in the series. For example, test form 1 would have five reading passages identified as A, B, C, D, E, while test form 2 would have five reading passages identified as D, E, F, G, H, and test form 3 would have five reading passages identified as G, H, I, J, K, etc. The passages in this series of test forms increased systematically in level of passage readability based on the Spache (1953) or the Dale-Chall (1948) readability formulas. This test series was designed to correspond with levels of student ability from grade 1 through grade 12, with approximately 135 students responding to one test form in the series. Thus, the entire Rasch calibration design for this test series required over 10,000 students. This large number of students provided confidence in the estimation of Rasch item difficulties and the subsequent chaining and scaling of the 200 reading passages on one difficulty scale.

It should be noted that the tests in this calibration series were constructed in multiple-choice cloze format and were designed to measure literal comprehension. One of the passages used in the calibration series is illustrated in Figure 1. Only nouns, verbs, adjectives, and adverbs were deleted from the passages and all item

---

Insert Figure 1 about here.

distractors were semantically implausible. Schuder, Kidder, and O'Reilly (Note 1) and Salter (Note 2) provide a more detailed description of this multiple-choice cloze test format.

In order to calibrate the 200 reading passages on the same difficulty scale, the following procedures were used:

1. Rasch item difficulties were calculated for each test form using the CALFIT program by Wright and Mead (Note 3).

2. The two identical passages on consecutive test forms provided a total of 20 items to be used in calculating the chaining constant for calibrating all test passages on the same difficulty scale.

3. The mean for the item difficulties on the first form that were also common to the second form was calculated and defined as $\bar{d}_{c1}$. The mean for the item difficulties on the second form that were common to the first form was calculated and defined as $\bar{d}_{c2}$.

4. The difference between $\bar{d}_{c1}$ and $\bar{d}_{c2}$ was used as a constant (c^t_{12}) that was then added to all of the item difficulties on the second form. That is,

$$\bar{d}_{c1} - \bar{d}_{c2} = c^t_{12}$$

and $d_{i2} + c^t_{12} = d'_{i2}$, meaning that all items on form 2 were increased by the value of the constant, c^t_{12}. The constant, c^t_{12}, at times was recalculated, and thus might go from c^1_{12} to c^2_{12} or c^3_{12}, depending on a number of iterations. The number of iterations performed on this constant (and for each constant between

consecutive forms, i.e., c^t_{23}, c^t_{34}, c^t_{45}, etc.) was based on t-values greater than 2 for individual items. The statistical significance of the difference between d_{i1} and d'_{i2} (the "bumped" difficulty of the common item on form 2) was determined by the following t test:

$$t = \frac{d_{i1} - d'_{i2}}{\sqrt{SEM^2_{i1} + SEM^2_{i2}}}$$

where SEM_{i1} is the standard error of measurement for item difficulty "i" on form 1 and SEM_{i2} is the standard error of measurement for the corresponding item difficulty on form 2.

5. If necessary, the chaining constant was recalculated by recomputing the mean of the common items on form 1 and then form 2, excluding items with significant absolute t-values. Item rejection was done one item at a time beginning with the "most" significant until a chaining constant was found based only on items with insignificant t-values.

6. An adjusted mean of the two final difficulties for each common item was used as the final estimate of the single item difficulty denoted as d^f_{ij}, that is,

$$d^f_{ij} = \frac{d_{i1}\, SEM_{i1} + d'_{i2}\, SEM}{SEM_{i1} + SEM_{i2}}$$

7. Passage difficulty (D) was the mean of the final single item difficulties in each passage following chaining. That is,

$$D = \frac{\sum_{i=1}^{n} d^f_{ij}}{n}$$

where n = the number of items in a passage.

8. Passage difficulties were then linearly transformed to D' scores on a scale beginning with 100. That is, D' = 10D + 102, where D' = Rasch passage difficulty.

The precision of the Rasch-based and the readability-based scaling techniques was determined as follows: (a) Passages with identical Spache or Dale-Chall values were selected and their Rasch difficulties inspected for equivalence. (b) When the Rasch difficulties were not equivalent, the grade level of the textual source of the passages was considered the criterion for passage difficulty. (c) The Rasch-based scaling was assumed to be more precise than the readability-based scaling, thus, passages with identical Spache or Dale-Chall readabilities would not have identical Rasch difficulties. In fact, the Rasch-based technique would scale passages in the same direction of difficulty as indicated by the grade level of the textual source of the passages. (d) The Rasch-based technique would be sensitive to syntactic and semantic complexities not reflected in the readability estimates for individual passages.

Results. Table 1 displays representative examples of Spache and Dale-Chall passage readabilities and Rasch passage difficulties for some of the 200 passages analyzed. Spache readability values ranged from 1.3 to 3.9, while Dale-Chall readabilities ranged from 4.50 to 8.45. Rasch passage difficulties ranged from 100.00 to 414.31.

Table 2 contains data on passages with identical Spache or Dale-Chall values, including the grade level of the textual source of the passage and the Rasch passage

difficulty. These data support the conclusion that the Rasch-based scaling provided more precise estimates of the inherent difficulty of the passages being studied; passages with identical readabilities had unequal Rasch passage difficulties and these Rasch passage difficulties scaled passages in the same direction of difficulty as indicated by the grade level of the textual source of the passages. In addition, the Rasch-based scaling seldom resulted in passages with identical difficulties.

Insert Tables 1 and 2 about here.

The sensitivity of the Rasch-based scaling is illustrated clearly in Figure 2. This figure contains two reading passages with identical Dale-Chall values. However, on the Rasch scale these two passages are 18.87 points apart. In fact, there were six additional passages scaled between 149.94 and 168.81 on the complete Rasch passage scale, therefore, these two passages must have structural or semantic complexities not represented in the two Dale-Chall variables, namely, percentage of hard words in a passage and average sentence length of a passage.

Insert Figure 2 about here.

A closer look at the syntactic and semantic structure of the two passages in Figure 2 reveals considerable structural differences between the passages. The "Ben Franklin" passage is obviously more complex than the "Charley Works"

passage. The first sentence in the "Ben Franklin" passage has an unusual phrase, namely, "he early learned," and the second and third sentences contain the figurative elements "just as" and "as if." Figurative elements are not found in the "Charley Works" passage. However, a more obvious difference in complexity is reflected in the high incidence of anaphora in the "Ben Franklin" passage, where every sentence contains anaphora, especially the third. These complexities are not reflected in the Dale-Chall formula scores, but the Rasch-based scaling of even these short reading passages is sensitive to variations in syntactic and semantic complexity.

Predicting Rasch Difficulty

Method

Correlation and regression analysis was used for three purposes: (a) to determine the degree of relationship between Rasch passage difficulty and Spache and Dale-Chall passage readability, (b) to predict Rasch passage difficulty from passage readability, and (c) to predict Rasch passage difficulty from the percentage of hard words in each passage and the average sentence length of each passage. The percentage of hard words in a Spache-level passage was the number of words in a passage not found on Clarence R. Stone's (1956) Revision of the Dale List of 769 Easy Words divided by the total number of words in the passage. For the Dale-Chall-level passages, the percentage of hard words in a passage was the number of words in a passage not found on the Dale List of 3,000 Familiar Words (Dale and Chall, 1948) divided by the total number of words in the passage.

The regression analyses, designed to determine the "fit" between Rasch passage difficulty and readability, were basically of two types: (a) models to determine the proportion of Rasch passage variance accounted for by linear, quadratic, cubic, and quartic readability effects, and (b) models to determine if Rasch passage difficulty was an exponential function of readability. With regard to the latter analysis, "Y" being defined as Rasch passage difficulty and "X" being defined as passage readability, two exponential functions were specified: (a) $Y = \alpha e^{\beta X}$, (b) $Y = \alpha X^{\beta}$. These exponential functions were analyzed using corresponding regression models in natural logarithmic form: (a) $\ln Y = \ln \alpha + \beta X$, (b) $\ln Y = \ln \alpha + \beta \ln X$.

Results. A scatter diagram showing the relationship between Rasch passage difficulty and Spache or Dale-Chall readability is provided in Figure 3. With reference to this figure, it is apparent that Rasch passage difficulty is positively related to readability, and, if reference is also made to Table 3, as expected, the zero order correlation between Rasch passage difficulty and readability is positive and high.

Insert Figure 3 and Table 3 about here.

However, some curvilinearity is suggested by the scatter diagram in Figure 3; thus, a series of regression analyses were completed to determine the "fit" between Rasch passage difficulty and readability.

Table 4 contains a summary of the first analysis of curvilinearity, namely, the proportion of Rasch variance accounted for by linear, quadratic, cubic, and quartic readability effects. It is apparent from Table 4 that the quadratic, cubic, and quartic readability effects contribute little to the total variance in Rasch passage difficulty. However, the latter generalization is somewhat less appropriate for the Spache-level passages. Nevertheless, the simple linear regression of readability on Rasch passage difficulty accounts for the largest proportion of Rasch variance for both Spache-level and Dale-Chall-level passages.

Table 5 contains the two log-based regressions for estimating the "fit" of the two exponential functions previously specified. Although the proportion of variance accounted for by each exponential function is substantial, it is not sufficient to negate the use of the simple linear regression model summarized in Table 4.

Insert Tables 4 and 5 about here.

Thus, Rasch passage difficulty can best be predicted from Spache or Dale-Chall readability using the following equations:

(A) Rasch passage difficulty = 74.56 + 18.07 (Spache readability)

(B) Rasch passage difficulty = -123.17 + 65.93 (Dale-Chall readability)

Note, the standard error of the regression coefficient was 1.17 for the Rasch-Spache regression and 1.08 for the Rasch-Dale-Chall regression, while the standard error of estimate was 6.606 for the Rasch-Spache regression and 14.094 for the Rasch-Dale-Chall regression.

The relationship between Rasch passage difficulty and readability having been specified, a third set of regression analyses was completed to determine the relationship between Rasch passage difficulty and the two variables used to calculate passage readability, namely, average sentence length and percent of hard words in a passage. Table 6 displays the results from these additional regressions. Thus, the two equations for predicting Rasch passage difficulty from average sentence length and percentage of hard words are as follows: (a) For Spache-level passages, Rasch passage difficulty = 90.69 + 2.3 (average sentence length) + 1.7 (percent hard words); (b) For Dale-Chall-level passages, Rasch passage difficulty = 118.74 + 3.07 (average sentence length) +10.57 (percent hard words). The standard error of estimate was 6.719 for the latter Rasch-Spache regression and 14.035 for the latter Rasch-Dale-Chall regression.

Insert Table 6 about here.

Predicting Grade Level of Text

Method

The 200 multiple-choice cloze exercises that were Rasch calibrated from least to most difficult constitute an instructional resource for determining the grade level of text a student can literally comprehend. For example, students entering grade 1 could be given a screening test composed of multiple-choice cloze exercises of increasing Rasch difficulty. The Rasch passage difficulties on

this screening test should be about 5 points apart. This screening test could be scored by using the difficulty of the last passage the student got 80 percent or better on as the student's level of competence. Then the difficulty of the last exercise passed could be used to estimate the grade level of text the student could literally comprehend. The latter estimate could be based on the regression of Rasch passage difficulty on grade level of textual source of passage.

Results. The regression of Rasch passage difficulty on grade level of textual source of passage was completed for the 200 passages being analyzed. From the latter analysis, the following regression equation was obtained: Predicted grade level of textual source of passage = -1.28 + 0.029 (Rasch passage difficulty); the standard error of estimate equaled 0.605. This equation can be used to estimate, within reason, the grade level of text a student can literally comprehend. For example, a grade 1 student could receive a screening test with 6 multiple-choice cloze exercises with the following Rasch passage difficulties: (a) 100.00, (b) 106.97, (c) 110.38, (d) 116.39, (e) 121.02, (f) 126.62. It is possible that a grade 1 student receiving such a test would score 80 percent or better on the first three exercises and below 80 percent on the last three exercises. The last passage this student scored 80 percent or better on had a Rasch difficulty of 110.38. This Rasch difficulty could be used to predict the grade level of text the student can literally comprehend by solving the following equation: Grade level of text = -1.28 + 0.029 (110.38), which

equals 1.92. Assuming a publisher's estimate of the grade level of a textbook is accurate, this student could probably literally comprehend the reading materials used during the second half of the first grade. However, the standard error of estimate for this prediction is quite large, i.e., 0.605, thus reducing confidence in this prediction of the grade level of text that students can literally comprehend.

Discussion

This study confirms the utility and precision of Rasch-based scaling of short reading passages and supports the conclusion that readability-based scaling is less precise than Rasch-based scaling. The relatively higher precision of Rasch-based scaling was observed even though Rasch passage difficulty and readability were highly correlated and linearly related. Comparisons between passages with identical readability estimates but unequal Rasch difficulties revealed that the Rasch scaling was sensitive to increased language complexity such as increased use of anaphora, figurative language, subordination, and abstraction.

Although Rasch passage difficulty can be predicted from Spache or Dale-Chall readability estimates, it is recommended that passages be scaled using the proposed Rasch-based techniques. This will prevent over or under-estimation of the inherent difficulty of the passage in question and will scale passage difficulty relative to the 200 "standard" Rasch-scaled passages.

The Rasch-based prediction of grade level of text a student can literally comprehend was disappointing. More precise prediction was expected. However,

publishers have not established calibration standards for textbooks or portions of textbooks; thus the error in the grade level prediction is probably due more to inconsistencies in the published grade level of the text than to inconsistencies in the Rasch passage calibrations. The scale of 200 Rasch-calibrated passages provides a new standard of reading complexity. In order to determine the relationship between textbook difficulty and this new standard of reading complexity, first it would be necessary to take coherent 80-word samples every tenth page (Barry, Note 5, p. 2) from the book in question. These 80-word samples would then be clozed in the multiple-choice format proposed (Hayford, et al., Note 4) and administered to an appropriate sample of students along with an equal number of passages from the "standard" scale of 200 passages. Following Rasch calibration and linking, the inherent difficulty of the text would be reflected in the Rasch passage difficulties of each newly sampled passage. This technique would establish the relative difficulty of the book in question, and it would indicate relative increases in difficulty from the beginning to the end of the book and unexpected or undesirable variations in the textual difficulty within the book.

The apparent language sensitivity of the proposed Rasch-based scaling technique supports the recommendation that the 200 scaled passages be used in studies designed to predict reading difficulty. For example, Olsen (Note 6) is using the 200 scaled passages as the dependent variable in a computerized analysis of textual complexity including predictor variables like proportion of

prepositions, transformations per sentence, grammatical groups per independent clause, and proportion of pronouns. This type of computerized text analysis could lead to better understanding of the syntactic and semantic factors contributing to reading difficulty.

Two additional areas of application for the 200 Rasch-scaled multiple-choice cloze exercises should be explored. First, parallel test forms assembled from the Rasch-scaled passages should be administered to the same students at two-month intervals for one school year in order to determine the test's sensitivity to reading instruction over time. Second, the accuracy of using the scaled multiple-choice cloze exercises for placement in reading materials of appropriate difficulty should be studied more carefully. Screening tests with exercises that rapidly increase in difficulty would help identify students who could literally comprehend textual materials within their particular range of reading skills. In fact, this type of screening test could even identify students having trouble with literal comprehension per se, an important finding in and of itself (Carroll, 1972, p. 3). These applications of the scaled multiple-choice cloze exercises will help to establish their instructional utility.

Reference Notes

1. Schuder, R. T., Kidder, S. J., & O'Reilly, R. P. Defining and measuring the literal comprehension of written discourse. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1976.

2. Salter, R. Test development notebook. Multiple-choice cloze exercises: Handbook. Albany, N. Y.: Bureau of School and Cultural Research, New York State Education Department, January 1977.

3. Wright, B. D., & Mead, R. J. CALFIT: Sample-free item calibration with a Rasch measurement model. (Research Memorandum Number 18) Chicago: The University of Chicago, Statistical Laboratory, Department of Education, March 1975.

4. Hayford, P. D., Schuder, R. T., & Kidder, S. J. Conventional and Rasch analysis of an unconventional test of literal comprehension. Report prepared for presession entitled Theory and Practice of Measurement with Rasch Models held during the meeting of the American Educational Research Association, San Francisco, CA, April 1976.

5. Barry, J. G. Computerized readability levels: their need and use. Paper presented at the meeting of the 6th World Congress on Reading, Singapore, August 1976.

6. Olsen, C. Personal communication, March 16, 1977.

References

Carroll, J. B. Defining language comprehension: Some speculations. In R. O. Freedle & J. B. Carroll (Eds.), Language comprehension and the acquisition of knowledge. New York. John Wiley, 1972, 1-29.

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20.

O'Reilly, R. P., & Streeter, R. E. Reports on the development and validation of a system for measuring literal comprehension in a multiple-choice cloze format. Journal of Reading Behavior, 9, 1977, in press.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen. Danish Institute for Educational Research, 1960.

Rasch, G. An individualistic approach to item analysis. In Lazarsfeld, P. & Henry (Eds.), Readings in mathematical social science. Chicago. Science Research Associates Inc., 1966, 89-107. (a)

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57. (b)

Spache, G. A. new readability formula for primary-grade reading materials. Elementary School Journal, 1953, 53, 410-413.

Spache, G. Good reading for poor readers. Champaign, Illinois. The Garrard Press, Publishers, 1960.

Stone, C. R. Measuring difficulty of primary reading material: A constructive criticism of Spache's measure. Elementary School Journal, 1956, 57, 36-41.

Wright, B. D., & Panchapakesan, N. A. procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Footnotes

The author would like to thank Ronald Streeter for completing the Rasch passage calibration, Benjamin Wright and Marshall Dahl for their helpful suggestions regarding the statistical procedures used, Charles Adams for assisting with test design and data acquisition, Eleanor Chamberlain for typing the manuscript, and Robert O'Reilly for initiating the investigation.

[1] The 200 Rasch-calibrated multiple-choice cloze passages were taken from the Test Development Notebook developed by the Bureau of School and Cultural Research, Division of Research, New York State Education Department.

[2] The complete set of 200 Rasch-calibrated multiple-choice cloze exercises is available from Steven J. Kidder, New York State Education Department, Division of Research, Room 481 EBA, Albany, New York 12234.

Table 1

Representative Passage Difficulty and Readability Estimates

Passage I.D.	Spache readability	Rasch D' calibration	Passage I.D.	Dale-Chall readability	Rasch D' calibration
01-01-08	1.4	104.40	03-07-27	4.64	149.94
01-01-06	1.3	102.88	03-07-28	4.50	154.71
01-01-07	1.4	100.00	06-09-03	5.01	194.68
01-02-29	1.8	101.28	04-09-32	5.00	199.84
01-02-34	1.5	101.42	06-12-26	5.84	254.81
01-02-38	1.5	103.94	06-12-25	5.90	267.04
01-02-41	1.8	106.97	06-14-33	6.38	306.79
01-02-20	1.7	111.23	07-14-17	6.31	302.19
01-02-21	1.5	106.23	09-18-17	7.47	372.15
01-02-22	1.7	108.08	09-18-18	7.48	364.36
02-04-38	2.9	113.68	10-21-25	8.00	389.17
03-05-05	3.2	118.49	10-21-28	8.17	388.30
04-06-06	3.9	160.42	10-22-32	8.45	396.15
04-06-04	3.9	164.13	10-22-29	8.26	414.31

Note. Of the 200 passages analyzed, 49 were Spache-level passages and 151 were Dale-Chall-level passages.

Table 2

Passage Readability, Grade Level of Textual Source of Passage, and Rasch Passage Difficulty

Passage readability	Grade level of source	Rasch passage difficulty
	Spache-level passages	
1.4	1	100.00
1.4	1	104.40
2.2	1	107.14
2.2	2	119.06
3.4	3	129.28
3.4	4	138.12
	Dale-Chall-level passages	
4.64	3	149.94
4.64	4	168.81
5.96	5	269.01
5.96	7	284.39
6.73	7	315.68
6.73	8	328.94
7.57	9	368.36
7.57	10	375.24

Table 3

Means, Standard Deviations, and Zero-order

Correlations by Passage Type

Descriptive data	Spache-level passage	Dale-Chall-level passage
Number of passages	49	151
Mean for readability	2.62	6.49
S.D. for readability	0.82	1.06
Mean for Rasch difficulty	121.99	304.64
S.D. for Rasch difficulty	16.15	71.57
Zero-order correlation between Rasch difficulty and readability		
Correlation coefficient	0.914	0.981

Table 4

Proportion of Rasch Variance Accounted for by

Regression of Linear, Quadratic, Cubic, and Quartic

Readability Effects on Rasch Passage Difficulty

Independent readability variables in equation	Spache-level passage	Dale-Chall-level passage
$X + X^2 + X^3 + X^4$	0.894	0.982
$X + X^2 + X^3$	0.865	0.982
$X + X^2$	0.864	0.978
X	0.836	0.962

Table 5

Proportion of Rasch Variance Accounted for

in Logarithmic Regression Models of

Readability on Rasch Passage Difficulty

Regression equation	Spache-level passage	Dale-Chall-level passage
$\ln Y = \ln \alpha + \beta X$	0.857	0.931
$\ln Y = \ln \alpha + \beta \ln X$	0.807	0.960

Table 6

Regression of Average Sentence Length and Percent Hard Words on Rasch Passage Difficulty

	Regression coefficient	Standard error	Intercept value	Multiple correlation coefficient
Spache-level passages				
Average sentence length	2.30	0.399	90.69	0.913
Percent hard words	1.70	0.202		
Dale-Chall-level passages				
Average sentence length	3.07	0.149	118.74	0.981
Percent hard words	10.57	0.204		

Figure Captions

Figure 1. Sample multiple-choice cloze exercise used in Rasch calibration test series.

Figure 2. Test passages with identical Dale-Chall readability estimates and different Rasch passage difficulty estimates.

Figure 3. Scatter diagram of Rasch passage difficulty estimates and Spache or Dale-Chall passage readability estimates.

BEN FRANKLIN LEARNS TO SWIM AND SAIL

Benjamin lived near the sea, and he _____ learned to swim and _____ . He never grew tired of _____ the wind carry the _____ over the water, just as it _____ his kite up into the _____ . One day, while he was _____, he fastened the _____ to himself as if it were a _____ and he were a _____ . It carried him gently over the water.

1. loudly
2. doubtfully
3. thickly
4. early
5. sharply

1. pasture
2. sky
3. flower
4. twinkle
5. ink

1. battle
2. sail
3. lurch
4. fry
5. feed

1. linking
2. flowering
3. flaring
4. grumbling
5. swimming

1. allowing
2. coaching
3. joining
4. watching
5. persuading

1. plane
2. strip
3. princess
4. kite
5. lump

1. cages
2. troops
3. boats
4. myths
5. blocks

1. god
2. strain
3. plain
4. sail
5. tusk

1. carried
2. tested
3. crowned
4. blocked
5. buttoned

1. weed
2. boat
3. bunny
4. coyote
5. handful

BEN FRANKLIN LEARNS TO SWIM AND SAIL

Benjamin lived near the sea, and he early learned to swim and sail. He never grew tired of watching the wind carry the boats over the water, just as it carried his kite up into the sky. One day, while he was swimming, he fastened the kite to himself as if it were a sail and he were a boat. It carried him gently over the water.

Rasch passage difficulty = 168.81

Dale-Chall = 4.64

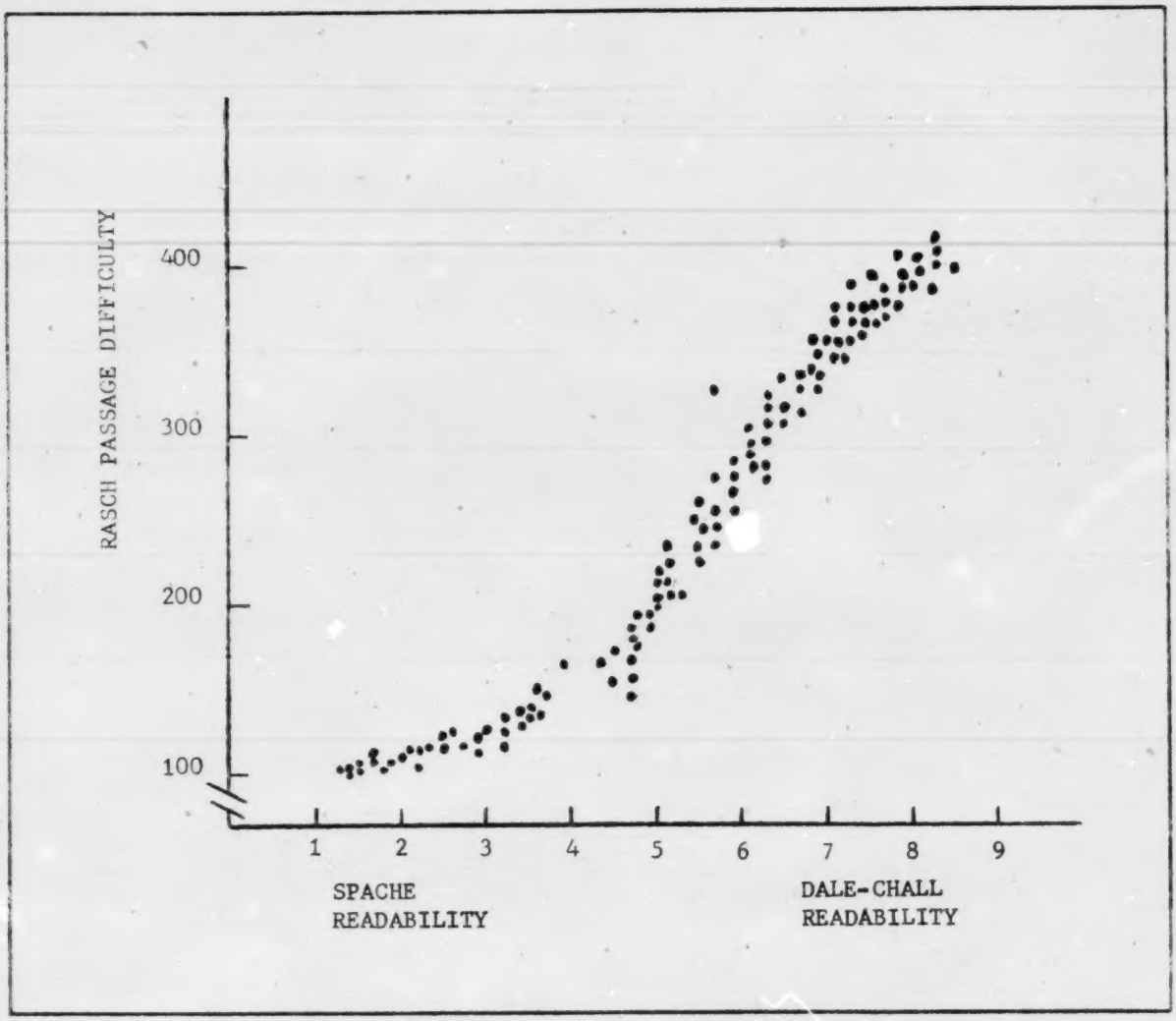
CHARLEY WORKS WITH THE CLAY

Mr. Sizemore and Charley spread newspaper on the floor from the middle of the room up to the door. Then Mr. Sizemore went back into the classroom where the other boys and girls were painting.

Alone in the room, Charley looked at the row of jars of clay standing on the shelf. He took down the jar containing the pink clay and went to work, rolling and rolling.

Rasch passage difficulty = 149.94

Dale-Chall = 4.64


RASCH PASSAGE DIFFICULTY
400
300
200
100
1
2
3
4
5
6
7
8
9
SPACHE
READABILITY
DALE-CHALL
READABILITY